UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES #DD-3

MEMORANDUM FOR Michael J. Longini
Chief, Decennial Systems and Contracts Management Office

Attention: Dennis W. Stoudt
Assistant Division Chief for Processing Systems

From: Howard Hogan
Chief, Decennial Statistical Studies Division

Subject: Specifications for the Identification of Coverage Edit Failures in Census 2000

I. INTRODUCTION

The purpose of this memorandum is to document the process for defining coverage edit failures for Census 2000. The coverage edit failures for Census 2000 will be for population count discrepancies and large household failures. Refer any questions on these specifications to Susan Ammenhauser on 301-457- 4236, or via cc:mail.

The Census 2000 coverage edit is a procedure to flag response records for followup. The universe for this edit is all records created from mail return short and long forms, Be Counted forms, and Internet responses. This universe does not include response records for enumerator-completed forms, telephone questionnaire assistance interviews, or any experimental form types.

A Headquarters edit of enumeration data will be done to identify records that have potential coverage errors and to identify records for which person data could not be reported on the mail return. Households whose records fail edit will be contacted by telephone. The telephone operation will be a computer-assisted telephone interviewing (CATI) followup that will be outsourced for Census 2000.

Section II defines the variables on the Data Capture System (DCS) 2000 Capture Output File used in specifying the coverage edits. It also provides definitions of computed values required for these edits.

The coverage edits specified in this memorandum assume that the Data Capture Audit and Resolution (DCAR) operation is implemented as specified in *DSSD Census 2000 Memorandum Series #K-2 (Revision 1)*, dated January 22,1999 (see attachment). It is assumed that data capture of each questionnaire is complete and that each response record with a potential count discrepancy has undergone a computer assisted review and resolution to validate person and roster name statuses. Section III defines criteria for determining which response records fail the coverage edits. Section IV of this document defines the tracking requirements for each edit failure. Section V documents other conditions that are applicable to coverage edits.

II. DEFINITIONS

A. Coverage Edit Variables

The majority of the variables used in the edits are defined during the DCAR process. (See the attachment for the complete definition of the DCAR process and DCAR edit variables.) Only records created from mail returns can have both the pre-audit and post-audit variables described below; the Internet and Be Counted form records will have only pre-audit variables because they do not go through the DCAR edits.

The critical variables required for the coverage edit are:

Variable Name	Definition
RN_POP	The value the respondent entered in question one (household size) on page one of all mail return questionnaires and Internet responses. This variable will reflect any corrections made to the data captured RN_POP value during DCAR (for mail returns). This value can be from 0 to 99. The RN_POP value may also be blank. Be Counted forms do not have this variable.
ROS_POP	The pre-audit DCAR count of valid roster names on long form mail return questionnaires. This value can be from 0 to 12. Short form mail returns, Be Counted forms and Internet responses do not have this variable.
VROS_POP	The post-audit count of valid roster names on long form mail return questionnaires after audit resolution is completed. VROS_POP is only filled for cases that go through audit. It is blank for cases that pass the DCAR edit. This value can be from 0 to 12. Short form mail returns, Be Counted forms and Internet responses do not have this variable.
IDP_POP	The pre-audit DCAR count of valid person panels for mail return, Be Counted form and Internet response records.

Variable Name	Definition
VDP_POP	The post-audit count of valid person panels. VDP_POP is only filled for cases that go through audit. It is blank for cases passing the DCAR edit. Be Counted forms and Internet responses do not have this variable.
CNT_POP	The pre-audit DCAR count of valid roster names on the continuation rosters of short form mail returns, Be Counted forms and Internet responses. This is the count of persons 7-12. This value can be from 0 to 6.
VCNT_POP	The post-audit count of valid roster names on continuation rosters of short form mail returns. VCNT_POP is filled only for cases that go through the audit. It is blank for cases passing the DCAR edit. This is the count of persons 7-12. This value can be from 0 to 6. Be Counted forms and Internet responses do not have this variable.
DCAREDIT	The status of the DCAR edit. Short and long form mail returns have this variable; Be Counted forms and Internet responses do not have this variable.
BCALLRES	Respondent's answer to question 4 on the Be Counted form. Indicates if form is completed for all persons in the household (whole household response) or only some household members (partial household response).

B. Definitions

The following defines a **valid person panel** (from the attached DCAR specs):

A VALID person panel is one that is Data Defined, is not coded CANCEL by KFI (key from image) and has not been determined by the DCAR Edit to be a possible DUPLICATE.

A **valid roster name** refers only to the first and last name fields. Together these two fields must have at least three legal characters for the names to be considered valid.

A **Data Defined Person record** is one where at least two of the following items have OCR- or KFI-accepted entries, or data fields in the cases of Internet records, that meet the following criteria:

- **Name**--Only the first and last name field are considered; together, these two fields must have at least three legal characters. [IMPORTANT: It is not required that a name be present for a person panel record to be considered data-defined or VALID].

- **Relationship**--A check box mark OR a write-in entry with at least one legal character. This item is not present for Person 1 records.

- **Sex**--A check box mark.

- **Age or Date of Birth**--Age OR Year of Birth OR Month and Day of Birth entry. [the assumption is that the range checks have been done on these four fields during previous operations so that DCAR need not perform them].

- **Hispanic Origin**--A check box mark OR a write-in entry with at least one legal character.

- **Race**--A check box mark OR a write-in entry with at least one legal character.

These DCAR definitions referenced in these specifications are used in determining if a valid name exists in valid person panels for persons 1 OR 2. They are also used in determining how to count valid roster names and valid person panels.

If persons 1 OR 2 meet the above definition of a valid person panel, then it must be determined if a valid name exists. For CEFU processing, a valid name is defined the same as the requirement for name (which is one of the items within the data defined person definition above).

C. Universe

The coverage edits are applied to the "normalized" file. Apply the coverage edits to all of the following response records:

* Short form mail returns D-1, D-1(S), D-1(T), D-1(C), D-1(K), D-1(V), D-1(UL)

* Long forms mail returns D-2, D-2(S), D-2(T), D-2(C), D-2(K), D-2(V) and D-2(UL)

* Internet responses

* Be Counted forms D-10, D-10(S), D-10(T), D-10(C), D-10(K), and D-10(V)

Response records must have a valid name for person 1 OR 2 in order to be sent to CEFU. An ID can be sent to CEFU only once. A flag will be maintained in order to identify the CEFU universe.

Be Counted form records must meet the following criteria in order to be eligible for the coverage edits.

* BCALLRES = 1 (form is for all persons in the household)

* have a MAF ID (that is, the record cannot have just a processing ID)

III. COVERAGE EDIT FAILURES

Response records that pass the DCAR edits (DCAREDIT=P) must use the pre-audit variables (ROS_POP, IDP_POP and CNT_POP) and RN_POP. Response records that fail the DCAR edit (DCAREDIT=F) must use the post-audit variables (VROS_POP, VDP_POP and VCNT_POP) and RN_POP. Note that Internet responses and Be Counted form records do not go through the DCAR edit; therefore, coverage edits for these records use pre-audit variables. The edits are defined below.

A. Short form mail returns that passed the DCAR edits, and Internet responses fail coverage edit based on any of the following conditions.

1. (RN_POP > 6)

2. (IDP_POP + CNT_POP > 6)

3. (RN_POP is blank **and** IDP_POP = 6 **and** CNT_POP = 0)

4. [(RN_POP is nonblank) **and** (RN_POP < IDP_POP)]

5. [(RN_POP is nonblank) **and** (RN_POP > IDP_POP)]

B. Long form mail returns that passed the DCAR edits fail coverage edit based on any of the following conditions.

1. (RN_POP > 6)

2. (ROS_POP > 6)

3. (RN_POP is blank **and** IDP_POP = 6 **and** ROS_POP = 0)

4. [(RN_POP is nonblank) **and** (RN_POP < IDP_POP)]

5. [(RN_POP is nonblank) **and** (RN_POP > IDP_POP)]

6. [(RN_POP is blank) **and** (IDP_POP < 6) **and** (ROS_POP < IDP_POP)]

7. [(RN_POP is blank) **and** (IDP_POP < 6) **and** (ROS_POP > IDP_POP)]

C. Short form mail returns that failed the DCAR edits fail coverage edit based on any of the following conditions.

1. (RN_POP > 6)

2. (VDP_POP + VCNT_POP > 6)

3. (RN_POP is blank **and** VDP_POP = 6 **and** VCNT_POP = 0)

4. [(RN_POP is nonblank) **and** (RN_POP < VDP_POP)]

5. [(RN_POP is nonblank) **and** (RN_POP > VDP_POP)]

D. Long form mail returns that failed the DCAR edits fail coverage edit based on any of the following conditions.

1. (RN_POP > 6)

2. (VROS_POP > 6)

3. (RN_POP is blank **and** VDP_POP = 6 **and** VROS_POP = 0)

4. [(RN_POP is nonblank) **and** (RN_POP < VDP_POP)]

5. [(RN_POP is nonblank) **and** (RN_POP > VDP_POP)]

6. [(RN_POP is blank) **and** (VDP_POP < 6) **and** (VROS_POP < VDP_POP)]

7. [(RN_POP is blank) **and** (VDP_POP < 6) **and** (VROS_POP > VDP_POP)]

E. Be Counted Forms can fail coverage edit based on the following condition.

IDP_POP + CNT_POP > 5 **and** there is at least one valid name on the continuation roster

IV. TRACKING REQUIREMENTS

A. Coverage Edits

Given the above coverage edit failure definitions, a response record can fail coverage edits for multiple reasons. Create flags that identify which edit(s) the record failed. The suggested flag variable name is in parenthesis following the edit name. The flag variable values are 0 = blank and 1 = failed edit.

1. Large Households (LHH)

These cases fail the large household edit. They are mail return or Internet households indicating a population count greater than 6, or Be Counted form households indicating a population count greater than 5.

a. Records pass DCAR edit, or do not have DCAR edit applied - use pre-audit variables

[(RN_POP > 6) **or** (IDP_POP + CNT_POP > 6)] - short form mail returns and Internet responses

or

([RN_POP > 6) **or** (ROS_POP > 6)] - long form mail returns

or

(IDP_POP + CNT_POP > 5 **and** there is at least one valid name on the continuation roster) - Be Counted forms.

b. Records fail DCAR edit - use post-audit variables

[(RN_POP > 6) **or** (VDP_POP + VCNT_POP > 6)] - short form mail returns

or

([RN_POP > 6) **or** (VROS_POP > 6)] - long form mail returns

2. Possible Large Household (P_LHH)

These cases are possible large households. They are mail returns or Internet responses with a blank response to question 1 and there are exactly six persons reported within the person panels. In addition, no one is reported on the roster or the continuation roster.

a. Records pass DCAR edit, or do not have DCAR edit applied- use pre-audit variables

[(RN_POP is blank **and** IDP_POP = 6 **and** CNT_POP = 0)] - short form mail returns and Internet responses

or

[(RN_POP is blank **and** IDP_POP = 6 **and** ROS_POP = 0)] - long form mail returns

b. Records fail DCAR edit - use post-audit variables

[(RN_POP is blank **and** VDP_POP = 6 **and** VCNT_POP = 0)] - short form mail returns

or

[(RN_POP is blank **and** VDP_POP = 6 **and** VROS_POP = 0)] - long form mail returns

3. Count Difference - High Data Defined Persons (H_DDP)

These cases are pop count discrepancies where there are more data defined persons than are reported in question 1 for short form and long form mail returns and Internet responses.

a. Records pass DCAR edit, or do not have DCAR edit applied - use pre-audit variables

RN_POP is nonblank **and** (RN_POP < IDP_POP) - short and long form mail returns and Internet responses

or

RN_POP is blank **and** (IDP_POP < 6) **and**
(ROS_POP < IDP_POP) - long form mail returns

b. Records fail DCAR edit - use post-audit variables

RN_POP is nonblank **and** (RN_POP < VDP_POP) - short and long form mail returns

or

RN_POP is blank **and** (VDP_POP < 6) **and**
(VROS_POP < VDP_POP) - long form mail returns

4. Count Difference - Low Data Defined Persons (L_DDP)

These cases are pop count discrepancies where there are fewer data defined persons than are reported in question 1 for short form and long form mail returns and Internet responses.

a. Records pass DCAR edit, or do not have DCAR edit applied - use pre-audit variables

RN_POP is nonblank **and** (RN_POP > IDP_POP) - short and long form mail returns and Internet responses

or

RN_POP is blank **and** (IDP_POP < 6) **and**
(ROS_POP > IDP_POP) - long form mail returns

b. Records fail DCAR edit - use post-audit variables

RN_POP is nonblank **and** (RN_POP > VDP_POP) - short and long form mail returns

or

RN_POP is blank **and** (VDP_POP < 6) **and** (VROS_POP < VDP_POP) - long form mail returns

B. Track the CEFU Universe

An ID can be sent to CEFU only once. Create a flag that indicates that an ID has been sent to CEFU. Maintain this flag for identification of the CEFU universe, to verify that an ID has already been sent (especially for those IDs that have multiple data capture records), and for evaluation purposes.

V. OTHER CONDITIONS

All questionnaires identified as blank mail return forms should not be recognized as legitimate mail returns. Blank mail returns are defined in DSSD Census 2000 Memorandum Series #K-3, dated March 8, 1999.

Attachment

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List, Coverage Edits Team Distribution List, and

R. Pennington (DSSD)
K. Zajac "
R. Dimitri "
E. Whitworth "
E. Katzoff (DSCMO)

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

January 22,1999

DSSD CENSUS 2000 MEMORANDUM SERIES #K-2 [Revision 1]

MEMORANDUM FOR Carol Van Horn
Chief, Decennial Management Division

From: Howard Hogan
Chief, Decennial Statistical Studies Division

Attention: William Norfolk
Decennial Management Division

Subject: Preliminary Analysis of the Dress Rehearsal Data Capture Audit and Resolution (DCAR) Edit and Review and its Redefinition for Census 2000 [Revision 1]

The Dress Rehearsal Experience

The DCAR Edit represents the only context data comparison run by the DCS 2000 system on response record content. The edit derives a count of persons from information on the response records interpreted from housing unit questionnaires, and compares this derived number to a count entry read from the questionnaire directly. Cases that fail the edit comparisons are subjected to a computer-assisted Audit Resolution operation where operators review the questionnaire images and the various statuses and counts involved in the DCAR Edit comparisons. They determine if the count inconsistency is the result of respondent or enumerator error, possibly confounded by a scanning capture system, and can be resolved by correcting the person record statuses assigned by the edit, or if the inconsistency in the counts is real and can only be resolved through personal contact with the household through a coverage followup operation.

The DCAR process instituted in the dress rehearsal by DCS 2000 was run only on mail return records. The capture edit as specified and programmed resulted in a much larger computer-assisted Audit Resolution review workload than had been expected, and it was obvious that an analysis of the DCAR dress rehearsal

experience needed to be done and the process rethought for the census. No reports on the number of mail return records failing the DCAR Edit or information on the types of changes made by the Audit Resolution review step were known at the end of the dress rehearsal process.

An evaluation of the DCAR process had not been formally planned. DSSD accepted the responsibility for analyzing and reporting on the dress rehearsal experience, and to explore expanding the capture edit and computer-assisted review to Be Counted Forms and enumerator return records. This analysis of the dress rehearsal DCAR/Audit results was undertaken under the purview of the DCAR working group chaired by Bill Norfolk (DMD) to determine the effectiveness of the dress rehearsal version of the process.

Although no counts of edit failures or overall workloads were available, both Lockheed Martin and NPC staff agreed that the lack of an answer to the respondent population count question (RN_POP) on the front of the short form mail return questionnaire was responsible for most of the Audit workload during dress rehearsal. NPC staff also reported that the RN_POP entry was often misread, and that both blank and non-blank entries reported by the system did not always reflect what the respondent had entered on the questionnaire. DSSD's initial focus, therefore, was on the RN_POP field, especially since its interpreted value or lack thereof dictated most of the dress rehearsal DCAR Edit failures.

The DCAR working group requested and received a log file from Lockheed Martin that allowed the DCAR failure rate for the dress rehearsal to be calculated, the reasons for edit failure to be estimated, and the changes in RN_POP made in the Audit Resolution operation to be quantified. The DCAR log files were merged with their matching records from DSCMO's normalized files to make this comparison as well as others.

The analysis showed that over 10 percent of short form mail return records from the Columbia and the Sacramento sites, and over 11 percent from the Menominee site failed the DCAR Edit and were reviewed by Audit Resolution. About 60 percent of all failures were a result of no entry or a non-numeric entry being captured for RN_POP. Blank RN_POP fields had been anecdotally blamed for the majority of DCAR failures, but the analysis also bore out observations by NPC staff that the RN_POP entry for many reviewed cases had not been accurately read. Preliminary analysis shows that over 13 percent of all RN_POP values associated with DCAR failures were changed in Audit Resolution after review of

the questionnaire image. Audit made an entry in about 7 percent of all the blank RN_POP fields and corrected the entry in over 26 percent of the non-blank numeric RN_POP entries after reviewing the question on the image.

Further analysis indicated that over 13 percent of all DCAR Edit short form failures required only review and corrected reentry of the RN_POP answer from the questionnaire image to bring the count comparison inconsistency to agreement, whereas 12 percent of the failing records required a more complex review of the person panels and roster names. The remaining response records continued to exhibit a person count inconsistency after the Audit Resolution process and, therefore, were candidates for a coverage followup operation.

The DCAR Edit requirement and specification presented here represent a concerted attempt, based on careful analysis of the dress rehearsal data available, to maximize the accuracy of the RN_POP variable and the FCU_POP variable (the population count entered by the enumerator after the NRFU interview) in the most efficient way. The types of cases found to have failed the edit predominantly because of inaccurate capture of the count fields should be treated as Key From Image (KFI)-type field corrections and not as full Audit cases that require a more costly and labor-intensive status review for possible erroneous person panels and roster names. Unlike in the dress rehearsal version of the edit, the Census 2000 DCAR Edit does not require that cases with blank RN_POP or FCU_POP fields fail if it can be shown that the OCR confidence thresholds and the KFI quality assurance on these two fields are tightened to capture and interpret them with 98.8 percent accuracy .[1]

A review of the questionnaires found to have unread RN_POP entries prior to Audit indicated that implementing an extended presence scan beyond the 2-digit segmented area when the initial OCR interpretation of the field was a blank might have allowed a significant number of these RN_POP fields be identified and

---

[1] The system failed to recognize that RN_POP was filled 0.5% of the time and recognized the presence of an entry but misinterpreted RN_POP 1.4% of the time. The total error on the RN_POP field is estimated to be 1.7%. This implies that the system achieved a threshold of 98.3% during dress rehearsal for this variable. We suggest that two confidence thresholds be defined. The first would stipulate how accurate OCR must be in recognizing the presence of an entry in RN_POP or in FCU_POP. OCR should be able to detect the presence of an entry 99.9% of the time. The second threshold would stipulate how accurately OCR must be in interpreting a detected entry. OCR should correctly interpret the value of these two variables 98.8% of the time.

captured by the KFI operation.

The requirements specified in this memorandum assume that this specified accuracy has been demonstrated by DCS 2000 using mail and enumerator questionnaires completed during the Dress Rehearsal enumeration. The table attached to this memorandum summarizes the results of the DSSD analysis, and provides the assumptions with which to arrive at Census 2000 workload estimates for the mail return portion of the Audit Resolution Count Check and Status Review operations. The same assumptions should be used for the enumerator returns at this time. A full report on the DCAR analysis will be issued by DSSD that will include more direct estimates for enumerator returns.

Based on the analysis, DSSD requires that DCS 2000 identify

- every RN_POP field and FCU_POP field that contains a non-numeric value, and every RN_POP field interpreted as blank by OCR but for which the presence of an entry in the immediate vicinity of the segmented area is detected,

- every RN_POP and FCU_POP numeric entry greater than the DCAR-derived value of TOT_POP for the return, and

- every RN_POP and FCU_POP numeric entry less than the DCAR-derived value of TOT_POP for the return.

Each of these situations are to be reviewed in a different manner:

- OCR must reject all RN_POP and FCU_POP fields containing non-numeric characters, and all RN_POP fields that are interpreted as blank but for which the presence of an entry in the immediate vicinity of the segmented area is detected. The KFI operation should review and correct these fields on all response records before the DCAR variables are created and the DCAR Edit is run.

- The DCAR Edit must identify all RN_POP and FCU_POP fields that are found to be greater than the derived TOT_POP value. These fields should be sent to an Audit Resolution Count Check operation to be reviewed and corrected before the final disposition of the response record, i.e, the creation and transmission of the T13 response record for the ID.

- The DCAR Edit must fail all response records for which RN_POP or FCU_POP is less than the derived TOT_POP value and send them to be reviewed by an Audit Resolution Status Review operation before the final disposition of the response record, ie., the creation and transmission of the T113 response record for the ID.

The size of the RN_POP and FCU_POP snippets presented to Audit Count Check should be of sufficient size that the keyer can see respondents' answers or notes written in the general area outside the small two-digit segmented write-in.

Reviewing the accuracy of RN_POP in this manner would have halved the number of dress rehearsal Audit Resolution cases requiring access to and review of the full questionnaire image and screens displaying the summary of DCAR statuses, as well as the program allowing status correction. Requiring only a KFI-type review and correction process on cases where the captured population count field value is greater that the DCAR-calculated total population should result in a significantly more efficient Audit operation and a barely measureable increase in the Coverage Followup workload (see Attachment A). RN_POP is one of the most important respondent entries on the census questionnaire. It drives not only the DCAR Edit but it plays a crucial role in identifying mail return households that must be contacted for coverage-related reasons. Both RN_POP and FCU_POP are critical in determining the final count of person counted in the census.

DSSD requires that counts of DCAR Edit failures by reason for failure by form type be kept by DCS 2000, in addition to the number and types of changes made in Audit Resolution . At the end of the census this data is to be made available to DSSD for operational analysis and evaluation.

The Requirements for the DCAR Edit and Audit Resolution in Census 2000

This requirement and specification attempts to use the same variable names that were employed in the dress rehearsal specification where appropriate.

Format Restriction on the Count Fields

OCR should reject all RN_POP and FCU_POP fields interpreted as having a non-numeric character in either digit of the field. The keyer should review the entry

and interpret, if possible, what the respondent entered. If the keyer cannot decode the entry to arrive at a numeric entry the field should be blanked. 'One' is to be entered '01', 'two' as '02', etc. All cases for which an alphabetic is still present when DCAR Edit is run must fail and be sent to Audit Resolution.

CANCEL Status Requirement on Person Panels and Roster Names

Preliminary analysis has shown that in dress rehearsal about 46 percent of all the short form DCAR failures for which RN_POP was filled were cases where the variable value was less than the count of persons derived from person panels and roster names. After reviewing the image, Audit Resolution changed the status on person panels or roster names on more than 60 percent of these cases such that RN_POP then agreed with the count of persons derived by the DCAR Edit. The number of edit failures will be decreased by identifying situations where data for an erroneous person panel or and erroneous roster name has been captured, and assigning them a CANCEL status. A variable referencing the first name field and one referencing the last name field for each person panel and each roster position should be set whenever a keyer blanks one of these fields. Respondents as well as enumerators often cross out the first and last name fields when they realize they have included a person on the questionnaire in error. The keyer, when presented with a crossed out name field, is instructed to blank the entry. The blanking action should be recorded in the applicable blanking variable. These variables are to be used by the DCAR Edit to identify person panels or the roster name positions as a CANCEL, and to set their statuses accordingly.[2]

Setting the CANCEL Status

If OCR detects the presence of an entry in the LAST NAME field AND if KFI blanks the field, THEN set the LAST NAME cancellation flag (PLN_KFIB or RL_KFIB) equal to 1; otherwise set the LAST NAME cancellation flag equal to 0.

If OCR detects the presence of an entry in the FIRST NAME field AND if KFI blanks the field, THEN set the FIRST NAME cancellation flag (PFN_KFIB or RF_KFIB) equal to 1; otherwise set the FIRST NAME cancellation flag equal to 0.

---

[2] The setting of the blanking variables does not apply to situations where names have been crossed out but other names have been entered in the same name write-in areas. The keyer must blank and leave the name field blank for the variable to be set.

Situations where person panels or roster names should be coded as CANCEL may also be identified during Status Review in Audit Resolution as a result of a DCAR Edit failure.

The DCAR Process versus the Edit Universe

The DCAR Edit will not be performed on records created by the KFP operation, particularly with 100% verification. It is assumed that the KFP procedures do not have keyers enter data or names that either the respondent or the enumerator have crossed out, or key information from a person panel that has been marked as filled in error. It should be obvious to those keying from the actual questionnaire when the respondent is indicating that information has been entered on the form by mistake. The purpose of DCAR is to attempt to identify and rectify situations that may result in erroneous population counts because no clerical review of responses is done in preparation for data capture. KFP records must, however, go through the DCAR process that sets the person panel and roster name status variables.

The DCAR process is performed on every response record created from data captured from a housing unit questionnaire. A status is to be determined for every person panel record associated with a housing unit response record. A status is also to be set on all roster name positions on response records created from questionnaires that have rosters, i.e., Mail Return and Be Counted Form records. The DCAR Edit itself is performed only on Mail Return and Enumerator Return--the records with RN_POP or FCU_POP values associated with them. All records that fail the DCAR Edit are reviewed in Audit Resolution by the simple Count Check or more rigorous Status Review process. Count Check requires only the operator view enlarged image snippets of the RN_POP or FCU_POP write-in areas and correct capture errors, while Status Review requires that entire person panel and roster name areas of the questionnaire images be perused by the operator to either verify or correct the various status set by the DCAR process.[3]

---

[3] DCAR statuses are set on response records created from the following stateside form types: D-1; D-1(UL); D-1(S); D-1(E); D-1(E)(SUPP); D-2; D2(UL); D-2(S); D-2(E); D-2(E)(SUPP); D-10; D-10(S). DCAR statuses are also set on the response records created from the following Puerto Rico form types: D-1(UL); D-1(UL)(S); D-1(E); D-1(E)(S); D-1(E)(SUPP); D-1(E)(SUPP)(S); D-2(UL); D-2(UL)(S); D-2(E); D-2(E)(S); D-2(E)(SUPP); D-2(E)(SUPP)(S); D-10; D-10(S). The DCAR Edit is run on records created from the following stateside form types: D-1; D-1(UL); D-1(S); D-1(E); D-2; D-2(UL); D-2(S); D-2(E). The DCAR Edit is also run on records created from the following Puerto Rico form types: D-1(UL); D-1(UL)(S); D-1(E); D-1(E)(S), D-2(UL); D-2(UL)(S); D-2(E); D-2(E)(S).

The DCAR Process Variables

Person Panel Statuses Defined:
A **VALID** person panel is one that is Data Defined, is not coded CANCEL either by KFI or, if an Enumerator Return panel, has its cancel box marked, and has not been determined by the DCAR Edit to be a possible DUPLICATE.

A Data Defined Person record is one where **at least two** of the following items have OCR- or KFI-accepted entries that meet the following criteria:

- **Name**--Only the first and last name field are considered; together, these two fields must have at least three legal characters. [IMPORTANT: It is not required that a name be present for a person panel record to be considered data-defined or VALID].

- **Relationship**--A check box mark OR a write-in entry with at least one legal character. This item is not present for Person 1 records.

- **Sex**--A check box mark

- **Age or Date of Birth**--Age OR Year of Birth OR Month and Day of Birth entry. [the assumption is that the range checks have been done on these four fields during previous operations so that DCAR need not perform them].

- **Hispanic Origin**--A check box mark OR a write-in entry with at least one legal character

- **Race**--A check box mark OR a write-in entry with at least one legal character

A **BLANK** person panel is one for which there are no entries at all on the person record, short or long. BLANK person panels have a P_STATUS of 'B'.

An **INVALID** person panel is one whose record does not meet the data defined person definition but has at least one item entry, i.e., person records with only one entry in any item (100% or sample), records with one 100% item entry and one or more sample item entries, and records no 100% item entries but any number of sample item entries. INVALID panels have a P_STATUS = 0.

A **DUPLICATE** person panel is a VALID panel that may possibly represent the same person as another person panel on the record. The first person panel in the pair of possible matches is considered to be the VALID panel, and the second person panel in the pair is the possible DUPLICATE. The first person panel is defined to be one with the lower person number. Matches between person panels will only be attempted for panels with first and last name entries AND complete dates of birth OR ages. The duplicate check is done only between VALID person panels on response records meeting specified conditions as defined in the DCAR Edit requirements. DUPLICATE panels have a P_STATUS = 2.

A DUPLICATE person panel is defined to be a panel for which

- the first name AND last name fields match exactly, and
- the age OR date of birth matches exactly

with another panel on the record.

A **CANCEL** person panel is one for which both the first and last name entries have been blanked in KFI, or for which one name field was blanked in KFI and OCR detected no entry in the other name field. For person panels on Enumerator Return records it is also defined to be panels for which the cancel box has been marked. CANCEL panels have a P_STATUS = 3. CANCEL person panels are not checked for possible duplication by the DCAR Edit.

Roster Position Statuses Defined:
A **VALID** roster name is one that meets the name definition described above for a Data Defined Person. VALID roster positions have an R_STATUS = 1.

An **INVALID** roster name is one that does not meet the name definition and was not coded CANCEL by the DCAR Edit. INVALID roster positions have an R_STATUS = 0.

A **BLANK** roster name is a roster position with no entry. BLANK roster positions have an R_STATUS of 'B'.

A **DUPLICATE** roster name is one with an exact first and last name match indicating that two roster positions may be duplicates of each other. The duplicate check is performed only between VALID roster names on response records meeting specified conditions defined by the DCAR Edit requirements. DUPLICATE roster names have an R_STATUS = 2.

A **CANCEL** roster name is one for which both the first and last name positions have been blanked in KFI, or for which one of the name fields was blanked in KFI and OCR detected no entry in the other name field. CANCEL roster positions have an R_STATUS = 3.

P_STATUS--the pre-Audit status of the person panel set by the DCAR Edit. It is used in the calculation of the person panel count IDP_POP.

R_STATUS--the pre-Audit status of the roster name position set by the DCAR Edit. It is used in the calculation of the roster name count CNT_POP and ROS_POP.

AP_STATUS--the post-Audit status of the person panel set only on response records that fail the DCAR Edit and are reviewed in Audit. It is used in the calculation of the person panel count VIDP_POP.

AR_STATUS--the post-Audit status of the roster name position set only on response records that fail the DCAR Edit and are reviewed in Audit. It is used in the calculation of the person panel count VCNT_POP and VROS_POP.

The Name Blanking Variables:
PLN_KFIB--when set for a person panel this variable indicates that the last name field has been blanked in KFI. It is used in the determination of a person panel CANCEL status.

PFN_KFIB--when set for a person panel this variable indicates that the first name field has been blanked in KFI. It is used in the determination of a CANCEL status for the person panel.

RL_KFIB--when set for a roster position this variable indicates that the last name field has been blanked in KFI. It is used in the determination of a CANCEL status for the roster position.

RF_KFIB--when set for a roster position this variable indicates that the first name field has been blanked in KFI. It is used in the determination of a CANCEL status for the roster position.

The Questionnaire and DCAR Edit Variables:
RN_POP--on the front of mail return and enumerator return forms, both short and

long. This variable represents the respondent entry or response to the inquiry "how many people were living or staying here on Census day?" This variable drives the DCAR Edit.

FCU_POP–on enumerator returns only; the Item B Pop count in the Interview Summary section on the back of the questionnaire. Along with RN_POP, this variable drives the DCAR Edit of enumerator returns.

ENUM_CANC–on enumerator returns only; when this box is marked the enumerator has "canceled" the person from the form and that panel is not counted. Panels with ENUM_CANC = 1 have a P_STATUS = 3.

IDP_POP--the pre-Audit DCAR count of VALID person panels, i.e, panels with P_STATUS = 1.

VDP_POP--the post-Audit count of VALID person panels. VDP_POP is filled ONLY for cases that go through Audit Resolution. It is blank for cases not failing the DCAR edit.

CNT_POP--the pre-Audit DCAR count of VALID roster names on the continuation rosters of short form mail returns and Be Counted forms, i.e., roster positions with R_STATUS = 1.

VCNT_POP--the post-Audit count of VALID roster names on continuation rosters after Audit Resolution is completed. VCNT_POP is filled ONLY for cases that go through Audit. It is blank for cases not failing the DCAR edit.

ROS_POP--the pre-Audit DCAR count of VALID roster names on long form mail returns, i.e., roster positions with R_STATUS = 1.

VROS_POP--the post-Audit count of VALID roster names on long form mail returns after Audit Resolution is completed. VROS_POP is filled ONLY for cases that go through Audit. It is blank for cases not failing the DCAR edit.

TOT_POP–the count derived by the pre-Audit DCAR Edit:

- For short form Mail Returns, TOT_POP = IDP_POP + CNT_POP.
- For long form Mail Returns, TOT_POP = IDP_POP OR ROS_POP, whichever is greater.
- For short form and long form Enumerator Returns, TOT_POP = IDP_POP.

- For Enumerator Supplemental returns, TOT_POP = IDP_POP.
- For Be Counted Forms, TOT_POP = IDP_POP + CNT_POP.

VTOT_POP--the post-Audit count derived after the review:
- For short form Mail Returns, VTOT_POP = VDP_POP + VCNT_POP.
- For long form Mail Returns, VTOT_POP = VDP_POP OR VROS_POP, whichever is greater.
- For short form and long form Enumerator Returns, VTOT_POP = VDP_POP.

DCAR_EDIT--the status of the DCAR edit. If the record passes the edit, DCAR_EDIT = P; if the record is eligible for Count Check DCAR_EDIT = C; if the record is eligible for Status Review, DCAR_EDIT = R.

Summary of values for P_STATUS, R_STATUS, AP_STATUS and AR_STATUS:
BLANK = B
INVALID = 0
VALID = 1
DUPLICATE = 2
CANCEL = 3

The DCAR Edit Defined

All DCAR Edit failures must be reviewed by Audit Resolution. The purpose of the DCAR Edit is to identify response records with count inconsistencies that may be the result of either OCR or KFI error or of respondent or enumerator error in filling the questionnaire--errors that, if not corrected, may result in needless and costly follow-up with respondents as a result of failing the Coverage Edit, or in errors in population count assumptions that result in the imputation of erroneous persons in the 100% edit and allocation process.

The OCR and KFI operations should be used to correct the values of RN_POP and FCU_POP when possible in a concerted attempt to limit the response records that fail the DCAR Edit to those that require a review of person panels and roster names on the image. However, whenever the DCAR TOT_POP count differs from the either the respondent's answer (RN_POP) or the enumerator's count (FCU_POP) these cases will be scrutinized further, either by checks made within the DCAR Edit, or by the Audit Resolution operation.

DCAR Edit for Short form Mail Returns:

Determine the number of VALID person panels (IDP_POP) and VALID roster names (CNT_POP). Calculate TOT_POP = INP_POP + CNT_POP.

If RN_POP = TOT_POP, the record PASSES

If RN_POP G.T. TOT_POP (including TOT_POP=0), the record FAILS. Check the accuracy of the RN_POP value by sending the record through the Audit Count Check operation, after which the record PASSES.

If RN_POP (including RN_POP blank and 0) L.T. TOT_POP, check for possible duplicates by comparing data between VALID person panels and between VALID roster names as defined above. If VALID panel(s) and/or roster name(s) change to DUPLICATE (s), recalculate TOT_POP based on new count of VALID panels and roster positions and compare RN_POP and the new TOT_POP.

If RN_POP is blank or 0, the record PASSES;
If RN_POP L.T. TOT_POP, the record FAILS. Check the accuracy of the person panel and roster name statuses by sending the record through the Audit Status Review operation, after which the record PASSES;
Otherwise the record PASSES.

DCAR Edit for Long form Mail Returns:

Determine the number of VALID person panels (IDP_POP) and VALID roster names (ROS_POP). Calculate TOT_POP as the larger of INP_POP and ROS_POP.

If RN_POP = TOT_POP, the record PASSES.

If RN_POP G.T. TOT_POP (including TOT_POP=0), check the accuracy of the RN_POP value by sending the record through the Audit Count Check operation, after which the record PASSES.

If RN_POP (including RN_POP blank and 0) L.T. TOT_POP, check for

possible duplicates by comparing data between VALID person panels and between VALID roster positions as defined above.
If VALID panel(s) and/or roster name(s) change to DUPLICATE (s), recalculate TOT_POP based on new count of VALID panels and roster positions and compare RN_POP and the new TOT_POP.

If RN_POP is blank or 0, the record PASSES;
If RN_POP L.T. TOT_POP, the record FAILS; check the accuracy of the person panel and roster name statuses by sending the record through the Audit Status Review operation, after which the record PASSES;
Otherwise the record PASSES.

DCAR Edit for Short and Long Form Enumerator Returns:

Determine the number of VALID person panels (IDP_POP). Set IDP_POP = TOT_POP.

If RN_POP = FCU_POP = TOT_POP, the record PASSES.

If RN_POP = FCU_POP AND G.T. TOT_POP, the record PASSES.

If RN_POP AND FCU_POP N.E. AND G.T. TOT_POP (including TOT_POP=0), check the accuracy of the RN_POP and FCU_POP values by sending them through the Audit Count Check, after which the record PASSES.

If RN_POP (including RN_POP blank and 0) OR FCU_POP (including FCU_POP blank and 0) L.T. TOT_POP, check for possible duplicates by comparing data between VALID person panels as defined above.
If VALID panel(s) change to DUPLICATE (s), recalculate TOT_POP based on the new count of VALID panels and compare RN_POP and FCU_POP with the new TOT_POP.

If RN_POP AND FCU_POP are blank or 0, the record PASSES;
If RN_POP OR FCU_POP are less than TOT_POP, the record FAILS; check the accuracy of the person panel and roster name statuses by sending the record through the Audit Status Review operation, after which the record PASSES;

Otherwise the record PASSES.

Be Counted Form records and Enumerator Supplemental records do not have either an RN_POP or FCU_POP field against which to judge the count of valid panel or roster records derived from data captured from the form. The applicable statuses for these records must be set by the DCAR process but the records are not subjected to the DCAR Edit and, therefore, are not candidates for review by Audit Resolution.

Audit Resolution

The purpose of Audit Resolution has not changed for Census 2000 from the dress rehearsal concept. However, to make the process more efficient the Audit process should be divided into two distinct operations--the Count Check, responsible only for correcting possible errors in the capture of the information provided in response to the RN_POP and the FCU_POP questionnaire items, and the Status Review, responsible for reviewing all possible reasons for the inconsistencies causing the response record to fail the DCAR Edit. It is also required for Census 2000 that all cases failing the DCAR Edit and processed through Audit Resolution be identified by a variable on the response record, identified here as DCAR_EDIT. In addition, response records that do not fail the DCAR Edit and therefore are not sent through the Audit Resolution process are not to have the post-Audit variables VDP_POP, VCNT_POP, and VROS_POP set. They are to remain blank, and not be set to the pre-Audit values for IDP_POP, CNT_POP, and ROS_POP. Also, two Audit variables, AP_STATUS and AR_STATUS, are to be added to each record so the actions by the Audit Resolution operator that result in changes to the person panel and roster name statuses will also be known, thus providing important information for monitoring and analysis that was lacking during and after the dress rehearsal.

Attachments

Summary of Dress Rehearsal DCAR Edit Results for Short Form Mail Returns

DCAR Edit Failures
Percent of Short Forms Captured

RN_POP Blank 6.5

6.7 % of Blank RN_POP had an entry (determined by Audit Resolution)

RN_POP NonNumeric 0.2

(17.6% of NonNumeric RN_POP were changed to a numeric in Audit Resolution)

RN_POP > TOT_POP 2.1

RN_POP < TOT_POP 2.0

DCAR Edit Failures with Numeric RN_POP
Percent of These Failures

	Status Before Audit Resolution	
Status After Audit Resolution	RN_POP > TOT_POP *(% of Column)*	RN_POP < TOT_POP *(% of Column)*
RN_POP=TOT_POP		
RN_POP Changed	21.8 *(41.9)*	3.0 *(6.2)*
TOT_POP or Both Changed	1.2 *(2.3)*	29.8 *(61.8)*
RN_POP ≠ TOT_POP	28.9 *(55.8)*	15.4 *(32.0)*

As this table shows, nearly all of the changes required for cases with RN_POP > TOT_POP were corrections to RN_POP. Only 2.3% of these cases resulted in corrections to the status of person panels or roster names, corrections that result in a change in the TOT_POP calculation, while 41.9 percent of these types of DCAR failures

consisted of only a correction in the capture of RN_POP, a change that can be made by reviewing just that field only. Based on this analysis, the records that fail the DCAR Edit because RN_POP > TOT_POP should be sent only to an Audit Count Check. The very small number of these cases that would have been resolved by the more labor-intensive status review will have only minimal impact on the workload for the Coverage Edit Followup if left to be resolved by that later process. It is not cost-effective to require that these cases be reviewed further by DCS2000.

SHORT AND LONG FORM MAIL RETURN DCAR EDIT LOGIC

if	RN_POP	equals	TOT_POP	then	Record PASSES
if	RN_POP	greater than	TOT_POP	then	Record FAILS; RN_POP is sent to Audit Count Check. Record then PASSES
if	RN_POP	less than	TOT_POP	then	Check for duplicates. Recalculate TOT_POP and compare RN_POP to new TOT_POP
After DUPLICATE Check					
if	RN_POP	equals	0 or BLANK	then	Record PASSES
if	RN_POP	equals	TOT_POP	then	Record PASSES
if	RN_POP	less than	TOT_POP	then	Record FAILS; person panel and roster name images are reviewed in Audit Status Review. Record then PASSES.

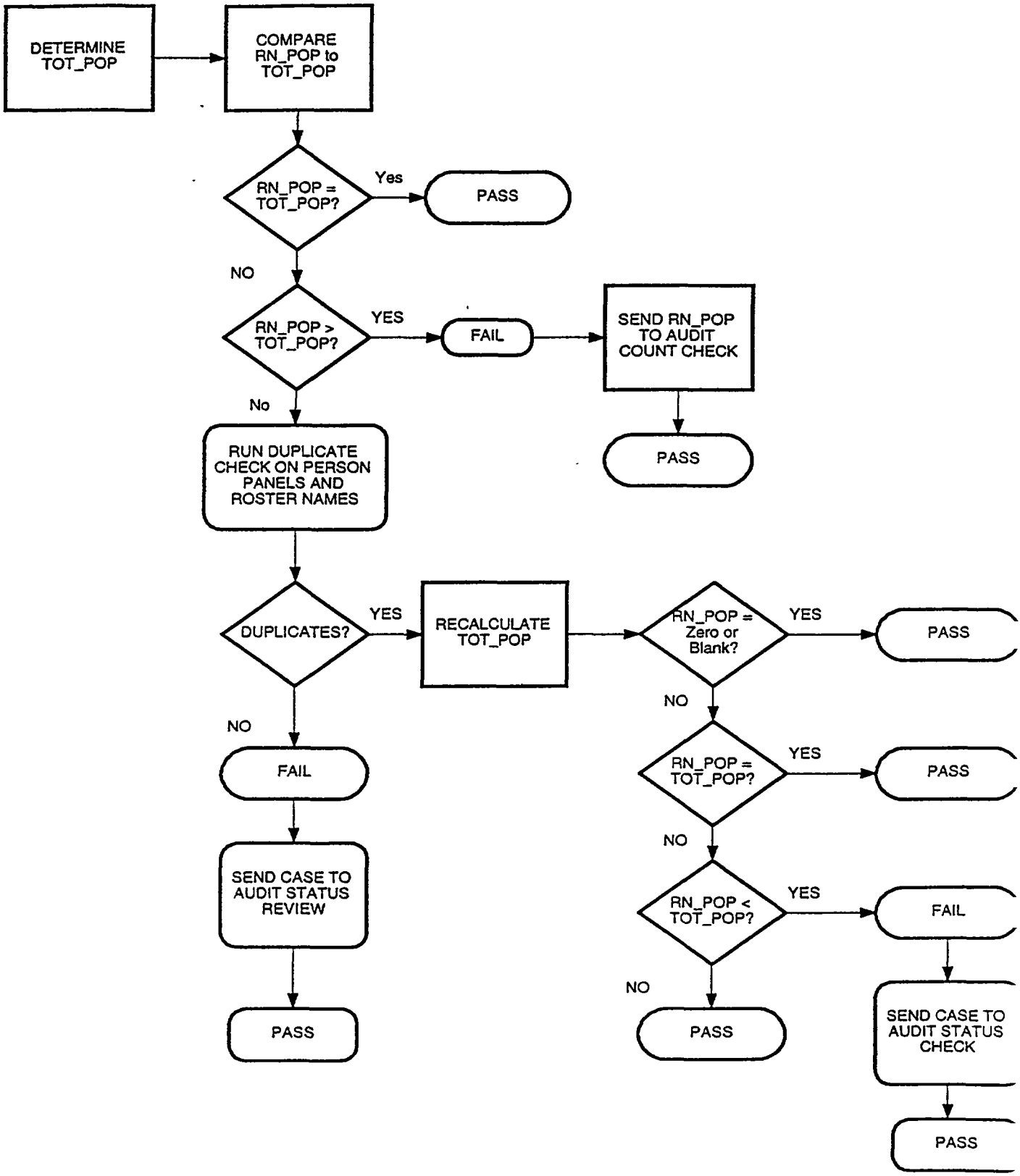
SHORT AND LONG FORM ENUMERATOR RETURN DCAR EDIT LOGIC

if	RN_POP	and	FCU_POP	equals	TOT_POP	Record PASSES
if	RN_POP	and	FCU_POP	are equal and greater than	TOT_POP	Record PASSES
if	RN_POP	and	FCU_POP	are not equal but greater than	TOT_POP	Record FAILS; RN_POP and FCU_POP are sent to Audit Count Check. Record then PASSES
if	RN_POP	or	FCU_POP	less than	TOT_POP	Check for duplicates. Recalculate TOT_POP and compare RN_POP and FCU_POP to new TOT_POP
After DUPLICATE Check						
if	RN_POP	and	FCU_POP	equal	0 or BLANK	Record PASSES
if	RN_POP	and	FCU_POP	equal	TOT_POP	Record PASSES
if	RN_POP	or	FCU_POP	less than	TOT_POP	Record FAILS; person panel and roster name images are reviewed in Audit Status Review. Record then PASSES

DSSD SPEC VARIABLE	DEFINITION	L/M OUTPUT VARIABLE
RN_POP	Respondent reported household population count from front cover of mail return and enum. questionnaire.	RNPOP
FCU_POP	Enumerator reported household population count from the Item B box of the Enumerator Questionnaire Interview Summary.	FCUPOP
P_STATUS	Person Status for person panels (result of DCAR status-setting phase) B=blank 0=Invalid 1=Valid 2=Duplicate 3=Cancel	PSTATUS
R_STATUS	Roster Person Status (also used for continuation roster persons) (result of DCAR status-setting phase) See P_STATUS for value assignments.	R##STAT (## = roster person number)
AP_STATUS	Person Status for person panels (result of DCAR Audit phase) See P_STATUS for value assignments.	APSTATUS
AR_STATUS	Roster Person Status (result of DCAR Audit phase) See P_STATUS for value assignments	ARSTATUS
PLN_KFIB	Person panel last name blanked by KFI 0 = OK 1=blanked	PLNKFIB
PFN_KFIB	Person panel first name blanked by KFI 0 = OK 1=blanked	PFNKFIB
RL_KFIB	Roster person last name blanked by KFI 0 = OK 1=blanked	R##LKFIB
RF_KFIB	Roster person first name blanked by KFI 0 = OK 1 = blanked	R##FKFIB
ENUM_CANC	Person cancel check box on SEQ's only: 0 = blank 1 = Cancel check box marked	ENUMCANC
IDP_POP	Pre-Audit valid person panel count (result of DCAR edit phase)	IDPPOP
VDP_POP	Post-Audit valid person panel count	VDPPOP
CNT_POP	Pre-Audit valid continuation roster name count (result of DCAR edit phase)	CNTPOP
VCNT_POP	Post-Audit valid continuation roster name count	VCNTPOP
ROS_POP	Pre-Audit valid roster name count (result of DCAR edit phase)	ROSPOP
VROS_POP	Post-Audit valid roster name count	VROSPOP
TOT_POP	Pre-Audit total person count	TOTPOP

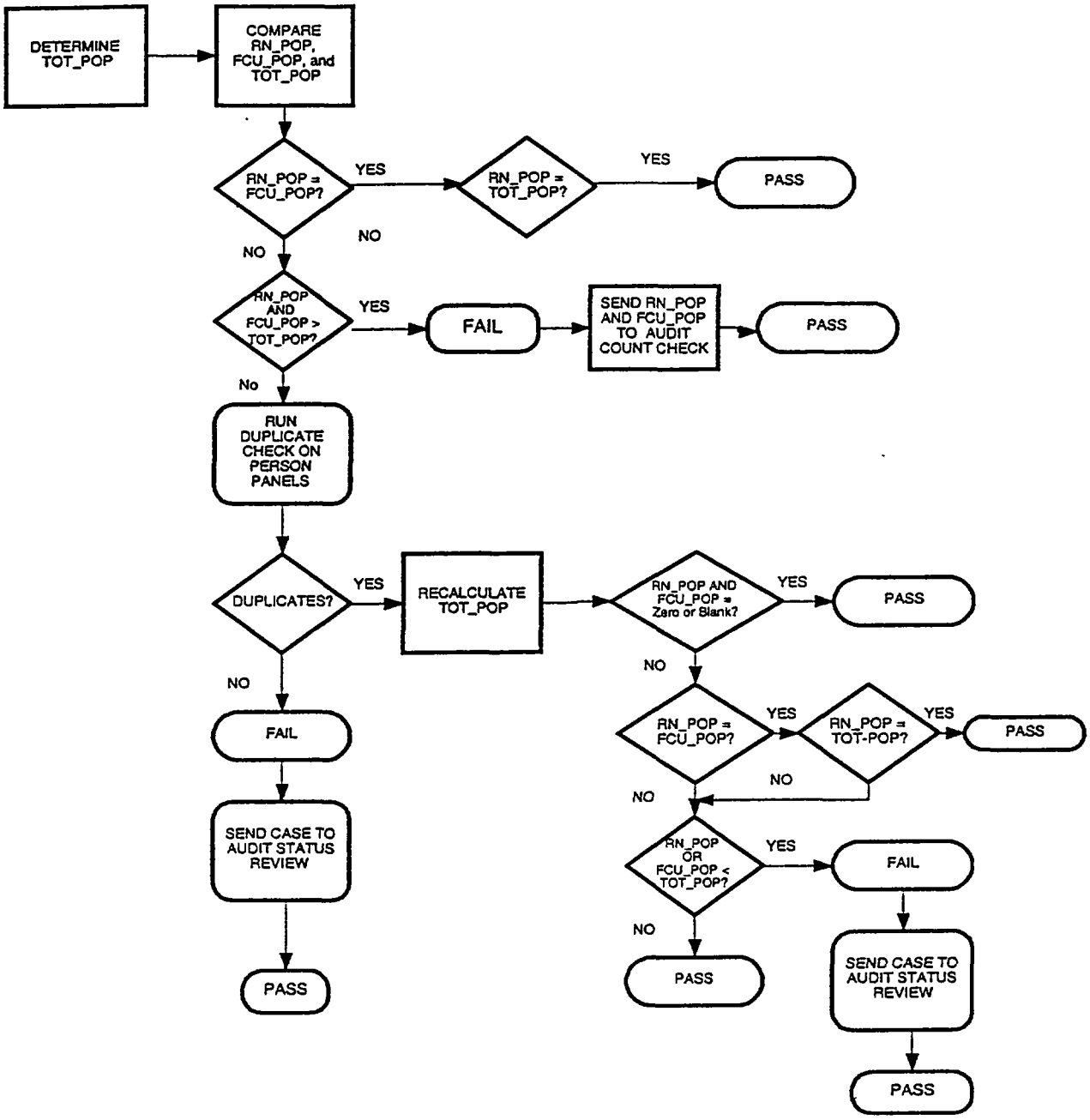
DSSD SPEC VARIABLE	DEFINITION	L/M OUTPUT VARIABLE
VTDT_POP	Post-Audit total person count	VTOTPOP
DCAR_EDIT	P=PASSED EDIT C=FAILED EDIT/Count Check R=FAILED EDIT/Status Review	DCAREDIT

DCAR EDIT FOR MAIL RETURN RESPONSE RECORDS



1/19/99

DCAR EDIT FOR ENUMERATOR RETURN RESPONSE RECORDS



1/19/99